

# REPORT



과목명		비모수통계자료분석
담당교수		최영훈 교수님
학과		응용통계학과
학년		3학년
학번		201452024
이름		박상희
제출일		2018년 10월 8일

## # 연습문제 5

개인별 인터뷰에 의한 여론조사 결과와 우편을 이용한 설문지 여론조사 결과가 차이점이 있는지를 살펴보고자 한다. 이를 위하여 세 명의 대통령 입후보자 A,B,C 에 대한 투표권자 100명의 표본을 추측하고, 임의로 40명은 개인별 인터뷰 조사를 나머지 60명은 우편 설문지 조사를 한 후 다음과 같이 투표권자의 반응 결과를 요약할 수 있었다. 과연 인터뷰나 설문지 여론결과에 있어서 세 후보자 사이의 차이가 존재하지 않는다고 할 수 있는가(다시 말하여 후보자들에 대한 선호도는 여론 조사 방법과 독립인가)?

조사방법	후보자			합계
	A	B	C	
인터뷰	18	6	16	40
설문지	20	15	25	60
합계	38	21	41	100

조사방법	후보자						합계
	A		B		C		
인터뷰	15.2		8.4		16.4		40
	18		6		16		
설문지	22.8		12.6		24.6		60
	20		15		25		
합계	38		21		41		100

### ① 가설

$H_0$  = 후보자들에 대한 선호도는 여론 조사 방법과 독립이다.

$H_1$  = 후보자들에 대한 선호도는 여론 조사 방법과 독립이 아니다.

### ② 검정통계량

$$T = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_{i=1}^r \sum_{j=1}^c \frac{O_{ij}^2}{E_{ij}} - N \quad (\text{단, } E_{ij} = \frac{r_i c_j}{N})$$

$$T = \frac{18^2}{15.2} + \frac{6^2}{8.4} + \frac{16^2}{16.4} + \frac{20^2}{22.8} + \frac{15^2}{12.6} + \frac{25^2}{24.6} - 100$$

$$T = 21.31 + 4.28 + 15.6 + 17.54 + 17.85 + 25.4 - 100 = 102.02 - 100 = 2.02$$

### ③ 통계적 검정

$T = 2.02 < \chi_{0.95, 2}^2 = 5.991$  이므로 귀무가설을 기각하지 못한다. 따라서 후보자들에 대한 선호도는 여론 조사 방법과 독립이라고 할 수 있다.

# 연습문제 13

교육문제 전문가들은 어머니의 교육수준이 초등학교 재학중인 자녀들에 대한 관심에 미치는 영향을 조사하고자 한다. 어머니의 교육수준을 측정하기 위한 수단으로는 어머니들이 가지고 있는 최종학위를 이용하였으며, 재학중인 자녀들에 대한 관심 정도를 측정하기 위한 수단으로는 어머니에 의한 재학중인 자녀들 학교의 자발적인 방문 횟수를 이용하였다. 확률표본 40명의 어머니들을 대상으로 조사한 아래 자료를 이용하여 어머니에 의한 학교방문의 수가 어머니 최종 학위 정도에 따라 차이가 존재하는지를 검정하여라.

학부모(어머니)의 최종학위				
국졸	중졸	고졸	대졸	대학원졸
2	4	0	4	4
3	6	3	3	2
1	2	5	2	5
3	2	4	9	2
4	3	2	5	
0	1	5	2	
5	0	2		
7	1	0		
0	5	6		
1	2	1		

	0 ~ 1	2	3 ~ 4	5 이상	
국졸	2.5 4	2.5 1	2.5 3	2.5 2	10
중졸	2.5 3	2.5 3	2.5 2	2.5 2	10
고졸	2.5 3	2.5 2	2.5 2	2.5 3	10
대졸 이상	2.5 0	2.5 4	2.5 3	2.5 3	10
	10	10	10	10	40

① 가설

$H_0$  = 어머니에 의한 학교방문의 수가 어머니 최종학위 정도에 따라 차이가 존재하지 않는다.

$H_1$  = 어머니에 의한 학교방문의 수가 어머니 최종학위 정도에 따라 차이가 존재한다.

② 검정통계량

$$T = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_{i=1}^r \sum_{j=1}^c \frac{O_{ij}^2}{E_{ij}} - N \quad (\text{단, } E_{ij} = \frac{r_i c_j}{N})$$

$$T = \frac{4^2}{2.5} + \frac{1^2}{2.5} + \frac{3^2}{2.5} + \frac{2^2}{2.5} + \frac{3^2}{2.5} + \frac{3^2}{2.5} + \frac{2^2}{2.5} + \frac{2^2}{2.5} + \frac{3^2}{2.5} + \frac{2^2}{2.5} + \frac{2^2}{2.5} + \frac{3^2}{2.5} + \frac{0^2}{2.5} + \frac{4^2}{2.5} + \frac{3^2}{2.5} + \frac{3^2}{2.5} - 40$$

$$T = 46.4 - 40 = 6.4$$

③ 통계적 검정

$T = 6.4 < \chi_{0.95, 9}^2 = 16.92$  이므로 귀무가설을 기각하지 못한다. 따라서 어머니에 의한 학교방문의 수가 어머니 최종 학위 정도에 따라 차이가 존재하지 않는다고 볼 수 있다.

# 연습문제 17

프로야구장 매표소에 입장권을 구입하기 위하여 늘어선 야구팬의 대기행렬이 성별에 상관없이 임의적인가를 알아보기 위하여 50명의 성별 순서에 따른 자료를 얻을 수가 있었다. Runs 검정을 하여라.

F F F M F M M M F F M F M M F M M  
 F M M M M F M F M M F M F F M M F  
 F F M F F M M F M F F F M M M F

=====

F	F	F	M	F	M	M	M	F	F	...	M	F	F	F	M	M	M	F
1	2	3	4	5	...	26	27	28	29									

① 가설

$H_0$  = 야구팬의 대기행렬이 성별에 상관없이 임의적이다.  
 $H_1$  = 야구팬의 대기행렬이 성별에 상관없이 임의적이지 않다.

② 검정통계량

$n_1$  = 남자(M)의 수     $n_2$  = 여자(F)의 수     $R$  = runs의 총수  
 $n_1 = 26$                        $n_2 = 24$                        $R = 29$

$$E(R) = \frac{2n_1n_2}{n_1 + n_2} + 1 = \frac{2 \times 26 \times 24}{26 + 24} + 1 = 25.96$$

$$Var(R) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)} = \frac{2 \times 26 \times 24 \times (2 \times 26 \times 24 - 26 - 24)}{(26 + 24)^2 \times (26 + 24 - 1)} = \frac{1,495,104}{122,500} = 12.2$$

$$Z = \frac{R - E(R)}{\sqrt{Var(R)}} = \frac{29 - 25.96}{\sqrt{12.2}} = \frac{3.04}{3.49} = 0.87 \sim N(0,1)$$

③ 통계적 검정

양측검정일 때 유의수준 5% 하에서  $Z = 0.87 < Z_{0.025} = 1.96$  이므로 귀무가설을 기각하지 못한다. 따라서 야구팬의 대기행렬이 성별에 상관없이 임의적이라고 할 수 있다.

# 연습문제 20

금리인하로 인하여 증권투자에 관심을 갖는 투자자가 증가하기 시작하였다. 증권 회사에서 투자자에게 권하는 투자예상종목 간에 차이가 있는지를 알아보기 위하여 증권투자를 시작한 10명의 고객을 임의로 추출하여 그들의 관심종목들에 대한 투자 여부를 분할하여 4곳의 증권회사와 상담, 투자하게 하였다. 6개월이 지난 후 이익을 보았다면 1, 반대로 손해를 보았다면 0으로 기록한 아래 자료를 얻을 수가 있었다. 이 자료를 통하여 4곳의 증권회사들 간의 예측은 차이가 있다고 말할 수 있는가?

증권회사	고객									
	1	2	3	4	5	6	7	8	9	10
A	0	0	1	1	1	0	0	0	1	1
B	0	0	0	1	1	0	1	1	1	1
C	1	1	0	1	1	1	1	1	0	0
D	1	0	1	1	0	1	1	0	0	1

	A	B	C	D	
1	6	6	6	6	24
	5	6	7	6	
0	4	4	4	4	16
	5	4	3	4	
	10	10	10	10	40

① 가설

$H_0$  = 4곳의 증권회사들 간의 예측은 차이가 없다.

$H_1$  = 4곳의 증권회사들 간의 예측은 차이가 있다.

② 검정통계량

$$T = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_{i=1}^r \sum_{j=1}^c \frac{O_{ij}^2}{E_{ij}} - N \quad (\text{단, } E_{ij} = \frac{r_i c_j}{N})$$

$$T = \frac{5^2 + 6^2 + 7^2 + 6^2}{6} + \frac{5^2 + 4^2 + 3^2 + 4^2}{4} - 40 = 24.3 + 16.5 - 40 = 40.8 - 40 = 0.8$$

$$T = 0.8 \sim \chi_{0.95, 3}^2$$

③ 통계적 검정

$T = 0.8 < \chi_{0.95, 3}^2 = 7.815$  이므로 귀무가설을 기각하지 못한다. 따라서 4곳의 증권회사들 간의 예측은 차이가 없다고 볼 수 있다.